Metabarcoding and Metagenomics 8: 295–320 (2024)
DOI: 10.3897/mbmg.8.128646

Research Article

Benchmarking the discrimination power of commonly used markers and amplicons in marine fish (e)DNA (meta)barcoding

João T. Fontes[1,2], Kazutaka Katoh[3], Rui Pires[1,2], Pedro Soares[1,2], Filipe O. Costa[1,2]

1 *Centre of Molecular and Environmental Biology (CBMA) and ARNET-Aquatic Research Network, Department of Biology, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal*
2 *Institute of Science and Innovation for Bio-Sustainability (IB-S), University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal*
3 *Research Institute for Microbial Diseases, Osaka University 3-1 Yamadaoka, Suita, Osaka 565-0871, Japan*

Corresponding author: Filipe O. Costa (fcosta@bio.uminho.pt)

Abstract

Environmental DNA (eDNA) metabarcoding is revolutionising the study of aquatic ecosystems, enabling high-throughput analysis of biodiversity with minimal disturbance. Despite its potential to support fisheries management, species identification and downstream analysis reliability are hindered by the lack of standardisation in DNA fragment choice. This study compares the species discrimination power of three markers used in marine fish (e)DNA (meta)barcoding – 12S rRNA, 16S rRNA and cytochrome oxidase subunit I (COI) – as well as two amplicons for each. We analysed sequences from NCBI GenBank for 10 orders of Actinopterygii (ray-finned fishes), including mitochondrial genomes. We assessed species discrimination by determining the percentage of monophyletic species in Neighbour-Joining trees and calculating congeneric divergences for two datasets: one with genomic regions extracted from mitochondrial genomes (771 species) and another with independent sequences for each region (3,879 species). Amongst (meta)barcoding amplicons in the mitochondrial genomes' dataset, the COI Folmer and Leray-Lobo regions had the highest discriminatory power, with 89.2% and 87.0% monophyletic species, respectively, while the 12S Teleo region had the lowest at 71.6%. Conversely, using independent sequences of these amplicons, Folmer and Leray-Lobo had the lowest percentages of monophyletic species, at 64.8% and 63.5%, respectively, while Actinopterygii 16S (Ac16S) had the highest at 83.0%. Species discrimination is influenced by the marker's evolutionary rate, fragment length, target fish order and the quality of reference sequence data. We recommend considering species discriminatory power differences for amplicon selection, especially for species-level identifications. We advise a standard multi-marker approach under certain scenarios, particularly when the presence of close congeneric species is expected.

Key words: Actinopterygii, environmental DNA, fisheries management, mitochondrial genome, molecular markers, species identification, taxonomic resolution

OPEN ACCESS

This article is part of:
Towards standardized molecular biodiversity monitoring
Edited by Teresita Porter, John Darling, Kelly Goodwin, Tiina Laamanen, Kristian Meissner, Toshifumi Minamoto

Academic editor: Kelly D. Goodwin
Received: 30 May 2024
Accepted: 23 August 2024
Published: 4 October 2024

Citation: Fontes JT, Katoh K, Pires R, Soares P, Costa FO (2024) Benchmarking the discrimination power of commonly used markers and amplicons in marine fish (e)DNA (meta)barcoding. Metabarcoding and Metagenomics 8: e128646. https://doi.org/10.3897/mbmg.8.128646

Copyright: © João T. Fontes et al.
This is an open access article distributed under terms of the Creative Commons Attribution License (Attribution 4.0 International – CC BY 4.0).

Introduction

The study of aquatic ecosystems is on the verge of a revolution due to emerging molecular tools, namely (e)DNA (meta)barcoding. With minimal disturbance to the environment, eDNA metabarcoding makes use of the genetic material that

is naturally shed by organisms to detect species using a high-throughput approach (Taberlet et al. 2012; Deiner et al. 2017). Considering the vastness and susceptibility of the marine ecosystem, it is clear how a high-throughput and non-invasive technique might be valuable for marine fish stock assessment, especially as a complement to traditional approaches (Stoeckle et al. 2021). However, despite the high potential held by this technique, taking the next step to fully integrate it into fisheries management decisions is highly dependent on several factors and challenges (Hansen et al. 2018; Gold et al. 2021a).

Fundamentally, the bioinformatics in taxonomic assignment during (e)DNA metabarcoding involves comparing high-throughput sequencing (HTS) reads of short-sized amplicons with reference sequences to identify species from an environmental or bulk sample (Taberlet et al. 2012; Mathon et al. 2021). Thus, the choice of the genomic region to be amplified and sequenced has multiple downstream implications (Othman et al. 2021; Xiong et al. 2022). For instance, the completeness of reference libraries is crucial for accurate (e)DNA (meta)barcoding-based species identification (Weigand et al. 2019; Gold et al. 2021b) and is currently one of the main obstacles in scaling-up the technique. Presently, this is especially challenging because different markers have substantially different coverages in terms of available public sequences and taxonomic groups (Machida et al. 2017; Hajibabaei et al. 2019; Marques et al. 2021). Another crucial, albeit under-assessed factor is the species-level discrimination capability of the marker and the respective amplicon used in an (e) DNA metabarcoding study (Polanco et al. 2021; Shu et al. 2021). This factor is of particular importance since different genetic markers are employed to target different taxonomic ranks (De Mandal et al. 2014) and some may not be as suitable for species-level identification.

Being the standard marker for DNA barcoding, COI is the backbone of animal species identification using molecular approaches (Hebert et al. 2003). Amongst commonly used DNA markers for (meta)barcoding, COI has the highest number of available animal reference sequences (Porter and Hajibabaei 2018; Weigand et al. 2019; Collins et al. 2021; Claver et al. 2023). In the case of marine fishes, COI has been extensively used to identify and delimit species, including cases with taxonomic ambiguities (Ward et al. 2005; Landi et al. 2014; Oliveira et al. 2016; Liu et al. 2022). However, primer bias and non-specific amplification of prokaryotes and non-target eukaryotes have been observed when using COI for fish metabarcoding, even when using primer pairs specifically targeting fishes (Collins et al. 2019). In response to this limitation, amplicons obtained from different mitochondrial DNA regions have become increasingly utilised in fish (e)DNA metabarcoding, most commonly amplicons within the 12S rRNA and 16S rRNA genes (Zhang et al. 2020; Xiong et al. 2022).

Both 12S and 16S have reported technical merits that make them useful for species detection (Miya et al. 2015; Berry et al. 2017; Zhang et al. 2020). In addition to showing higher specificity for the target group than COI (Collins et al. 2019), 12S and 16S generally comprise shorter amplicons (Valentini et al. 2016; Berry et al. 2017) and are consequently more likely to be preserved in environmental samples (Liu et al. 2020; Zhang et al. 2020). Despite their broad use in seawater eDNA-based monitoring of marine fish fauna, comprehensive assessment of the species-level discrimination capability of 12S and 16S and their most used amplicons remains wanting, particularly compared to

COI, where the existing extensive data have shown high discrimination power in fish species. Indeed, although several studies have analysed the discriminatory power of these markers, they were usually limited in scope, seldom directly comparing the three markers and respective amplicons and focusing on either a relatively small number of species or on a regional fish fauna (e.g. Collins et al. (2019); Polanco et al. (2021); Shu et al. (2021)).

Challenges for this type of analysis include the fragmented nature of DNA sequence data in public repositories, particularly for 12S and 16S, since there is no standardised region established for these markers, unlike the COI barcode region (Hebert et al. 2003). This implies that there are large imbalances in the species coverage of either the full-gene sequence data or their most common amplicons, which originate from different regions of the gene. Further to this challenge, compared to COI, non-protein-coding genes have considerable variation in size and sequence alignments are more prone to the occurrence of indels (Medina and Walsh 2000), thereby making it difficult to mine data and map these within the gene. Finally, as there is currently no global system in place to validate the quality and taxonomic congruence of sequence data available in public repositories, there may be inaccurate taxonomic assignments (Arranz et al. 2020; Fontes et al. 2021; Radulovici et al. 2021; Claver et al. 2023; Lavrador et al. 2023) that will compromise at least in part this type of analysis.

In this study, our goal was to conduct a comprehensive assessment of the capacity of commonly used markers in fish (e)DNA (meta)barcoding — 12S, 16S and COI — and six of their respective amplicons, to discriminate between species across 10 different orders of commercially and ecologically relevant marine Actinopterygii, thereby benchmarking their level of taxonomic resolution for (e)DNA-based studies. We utilised the concept of monophyly as the basis for determining species discriminatory power as it simplifies the quantification of taxonomic resolution by examining the congruence of sequences within specific clades.

Our analyses were conducted using two distinct datasets comprising publicly available sequences from NCBI GenBank: a) a dataset of sequences for the three markers extracted from mitochondrial genomes, thereby allowing direct comparison of discrimination ability between markers and amplicons for the same individuals and b) a dataset including independent sequences of the three full-gene markers, as well as their respective amplicons, allowing a comprehensive analysis of the discrimination capacity, based on the extant data. In addition, we calculated and compared each fragment's congeneric divergence to further grasp the potential reasons behind differences in discrimination capability.

This comprehensive evaluation highlights the critical need for standardisation in the selection and application of (e)DNA metabarcoding markers in marine fish studies. Establishing a standardised approach to marker choice, including their demonstrated taxonomic resolution and discrimination capability as criteria, will ensure the reliability and comparability of biodiversity assessments across different marine ecosystems. By benchmarking the effectiveness of these markers and their respective amplicons, our study supports the development of consensus guidelines that can enhance the accuracy of species-level identifications.

Methods

Data mining

We mined 12S, 16S and COI sequences for Actinopterygii from NCBI GenBank (Sayers et al. 2021), including partial and complete mitochondrial genomes, in September of 2022. To ensure comprehensive retrieval, we combined two approaches. First, we used a conventional search that checks GenBank accessions' annotated descriptions and titles (search terms are available in Suppl. material 1). Due to the non-uniform syntax across records, this approach may miss relevant data. Thus, we also used NCBI BLAST (Johnson et al. 2008) with a full-length 12S, 16S or COI sequence as a query, performing "BLASTn" searches for each order of Actinopterygii listed in the World Register of Marine Species (WoRMS Editorial Board 2023) and the NCBI Taxonomy database (Schoch et al. 2020). We downloaded all matching records, regardless of sequence size.

Using R (R Core Team 2023), we excluded records not identified at the species level, unverified records (i.e. see Suppl. material 2) and those with ambiguous species names (e.g., sp., cf. etc.). In addition, we removed sequences categorised as "COI-like" (Buhay 2009) from the COI dataset. Resorting to the *worms* R package (Holstein 2018), we retained only records belonging to marine species according to WoRMS habitat assignments. Subsequently, we used the taxonomic information in WoRMS to rectify unaccepted or synonym species names and standardise the taxonomy for downstream analyses.

To diminish the impact of indels, particularly in the cases of 12S and 16S regions and to obtain more reliable multiple sequence alignments, we made subsets of each marker's datasets into five groups of fishes. Each group comprised a pair of closely-related orders that were well-represented, meaning they had a substantial number of sequences, species and sequences per species. These pairings were selected according to the phylogeny reconstructed by Rabosky et al. (2018); (https://fishtreeoflife.org/taxonomy). The groups were as follows: Clupeiformes and Alepocephaliformes; Gobiiformes and Kurtiformes; Perciformes and Scorpaeniformes (although Scorpaeniformes are classified within Perciformes according to the WoRMS taxonomy, we have separated them here for clarity); Pleuronectiformes and Carangiformes (including Carangaria *incertae sedis*); and Scombriformes and Syngnathiformes. Hereafter, the following abbreviations will be used to refer to the groups: CA, GK, PS, PC and SS.

Extraction of target regions

Using the datasets obtained for each group of marine fishes and each marker, we extracted the following genomic regions (Fig. 1): complete 12S rRNA gene, MiFish-U region (Miya et al. 2015), Teleo region (Valentini et al. 2016), complete 16S rRNA gene, Berry-Fish region (Berry et al. 2017), Ac16S region (Evans et al. 2016), complete COI gene, COI DNA barcode Folmer region (Folmer et al. 1994; Hebert et al. 2003) and the Leray-Lobo region (Leray et al. 2013; Lobo et al. 2013). For brevity, the full-length regions of 12S, 16S and COI will be referred to as full 12S, full 16S and full COI, respectively and the Berry-Fish region as Berry. We selected two amplicons for each marker, targeting different regions within the full-length gene where possible. The widespread use of these amplicons (Xiong et al. 2022) was also a key criterion in our selection process.

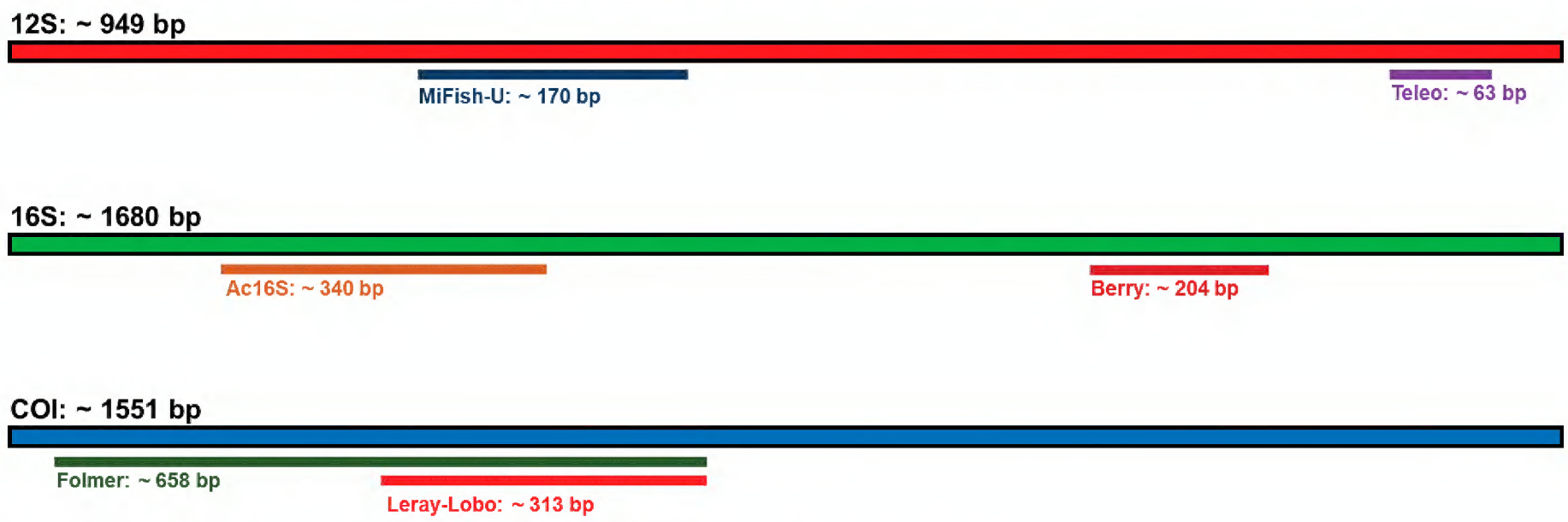

Figure 1. Representation of the nine genomic regions analysed in this study.

To extract the regions from the mined sequences, we used a MAFFT (Katoh et al. 2002) calculation (https://mafft.cbrc.jp/alignment/server/specificregion-last.html) employing the LAST algorithm (https://gitlab.com/mcfrith/last; Kiełbasa et al. (2011)) to find and locally align a query DNA region to a reference dataset, followed by multiple sequence alignment of the extracted sequences. For local alignment, we exclusively extracted sequences covering a minimum of 80% of the query region, as lower coverage sequences would be less viable during taxonomic assignment in (e)DNA metabarcoding analyses.

For consistency, we used query sequences from records containing all target regions, hereby focusing on complete or nearly complete mitochondrial genomes. For each pair of fish orders, we used one query sequence from each of the better-represented families in that group's mitochondrial genome datasets (i.e. families representing at least 10% of the dataset).

A custom Python script identified the primer annealing regions for the 12S and 16S amplicons to obtain the query sequences. For the full COI, 12S and 16S gene regions, which are readily annotated in the GenBank mitochondrial genome records, we downloaded sequences directly from the respective records' pages. For the Folmer region, since it usually lacks indels, we used ClustalW (Thompson et al. 1994) implemented in MEGA 11 (Tamura et al. 2021) to align the mitochondrial genomes to a reference barcode Folmer region (658 bp) sequence and trimmed the alignment. Finally, for the Leray-Lobo region, we trimmed the previous alignment to obtain the respective 3' 313 bp segment. For each dataset (i.e. one including sequences extracted from mitochondrial genomes and another including independent sequences for each genomic region), we obtained a total of 45 multiple sequence alignments, one per group of fishes and genomic region.

Data analysis

To determine and compare species-level discrimination power between regions, we solely used GenBank records where all target regions in Fig. 1 could be locally aligned (i.e. mitochondrial genome sequences). For each group of marine fishes, all target region datasets comprised sequences originating from the same GenBank record and, consequently, the same individual organism.

We realigned the sequences using MAFFT and proceeded to construct Neighbour-Joining trees in MEGA 11, with 1,000 bootstrap replicates. To ensure

consistency and estimate reliable trees for shorter regions (e.g. Teleo, MiFish-U or Berry), we used the uncorrected pairwise distance as the substitution model, with pairwise deletion for missing sites.

After obtaining the trees, we used the *MonoPhy* (Schwery and O'Meara 2016) R package to identify monophyletic and non-monophyletic species, calculating the percentage of monophyletic species for each region and group of fishes as a metric for species discrimination capability. Each Neighbour-Joining tree included an outgroup sequence from a different taxon within Actinopterygii, since the tree needs to be rooted for the R function to work. Singletons (i.e. species with only one sequence in the alignment) were classified as monophyletic unless they were intruders (i.e. sequences preventing another species from being monophyletic), in which case they were considered non-monophyletic. Preliminary analyses showed negligible differences in terms of the percentage of monophyletic species when testing other substitution models and tree-estimating methods.

For each genus, we used the *msa* R package (Bodenhofer et al. 2015) to iteratively realign the sequences with ClustalW and the *ape* R package (Paradis and Schliep 2019) to calculate the uncorrected pairwise congeneric distances and their averages. We realigned the sequences because this analysis was limited to genera comprising at least two species in the dataset.

We then performed the same analysis on the complete datasets, using sequences locally aligned for each independent genomic region. Due to computational restraints imposed by larger datasets (e.g. the Folmer region dataset for PS comprised over 15,000 sequences), we kept only two duplicate haplotypes per species for tree construction. We kept two duplicates instead of one to allow singletons the possibility of becoming intruders for species with only one haplotype, but more than one sequence. For calculating congeneric divergence values, we removed all duplicate haplotypes within each species. Preliminary analyses without duplicate removal showed no differences in results.

For the complete datasets, since they are not directly comparable between genomic regions, we calculated the percentage of monophyletic species and identified the reasons for non-monophyly: i) the percentage of non-monophyletic species sharing a haplotype with at least one sequence from another species, ii) the percentage of non-monophyletic species with at least one intruder, iii) the percentage of non-monophyletic species with at least one outlier (i.e. a sequence from a species clustering outside its species' clade) and iv) the percentage of non-monophyletic species with both an intruder and an outlier. These percentages were calculated excluding singletons, as they cannot have intruders or outliers, though they can become either. The R and Python scripts used in this study are available at https://github.com/tadeu95/Species_Discrimination. Additional methodology details are provided in Suppl. material 2 for clarification.

Results

The number of sequences and species that were locally aligned to the reference records mined from GenBank are represented in Table 1 and their GenBank accession numbers are provided in Suppl. material 3. Overall, across the five groups of fishes (per group can be observed in Suppl. material 4), the full 16S gene region had the lowest number of locally aligned sequences with 2,845,

while the full COI gene region had the lowest number of species with 1,064. On the other hand, the Leray-Lobo region had the highest number of sequences and species, respectively 49,757 and 3,314.

Species discrimination analysis using mitochondrial genomes

The dataset of mitochondrial genomes comprised 2,094 sequences across 771 species. The percentage of monophyletic species for the overall dataset, as well as for each group of fishes and genomic region, is shown in Fig. 2. The percentages varied from 59.4% of monophyletic species for the Teleo region in the PS dataset, to a maximum of 94.7% for the full COI region in the PC dataset. Amongst the metabarcoding amplicons analysed, the Leray-Lobo region had the highest percentage of monophyletic species in all groups. Teleo had the lowest percentage of monophyletic species in CA, PS and SS, while MiFish-U had the lowest percentage in GK and PC (Fig. 2). The average congeneric pairwise distance for each genus with at least two different species is represented in the box plots of Fig. 3, with detailed values for all genera analysed in the mitochondrial genomes' dataset available in Suppl. material 5. The overall number of observed haplotypes within each genomic region analysed is represented in Table 2, with the Teleo region having the lowest number of haplotypes (720) and the full COI region having the highest (1,276). The scatter plots in Fig. 4 show the congeneric pairwise distance values for six pairs of genomic regions.

Table 1. Number of sequences and species locally aligned using MAFFT for the five groups of marine fishes, for each genomic region analysed.

Region	Sequences	Species
Berry	10,482	2,061
Ac16S	4,078	1,521
Full 16S	2,845	1,069
Teleo	4,136	1,423
MiFish-U	7,780	2,117
Full 12S	3,127	1,104
Leray-Lobo	49,757	3,314
Folmer	48,331	3,310
Full COI	4,168	1,064

Table 2. Number of haplotypes observed in the mitochondrial genome dataset for each of the nine genomic regions analysed. Each dataset comprised 2,094 mitogenomes for 771 species.

Region	No. of Haplotypes
Berry	903
Ac16S	914
Full 16S	1,239
Teleo	720
MiFish-U	820
Full 12S	1,074
Leray-Lobo	981
Folmer	1,096
Full COI	1,276

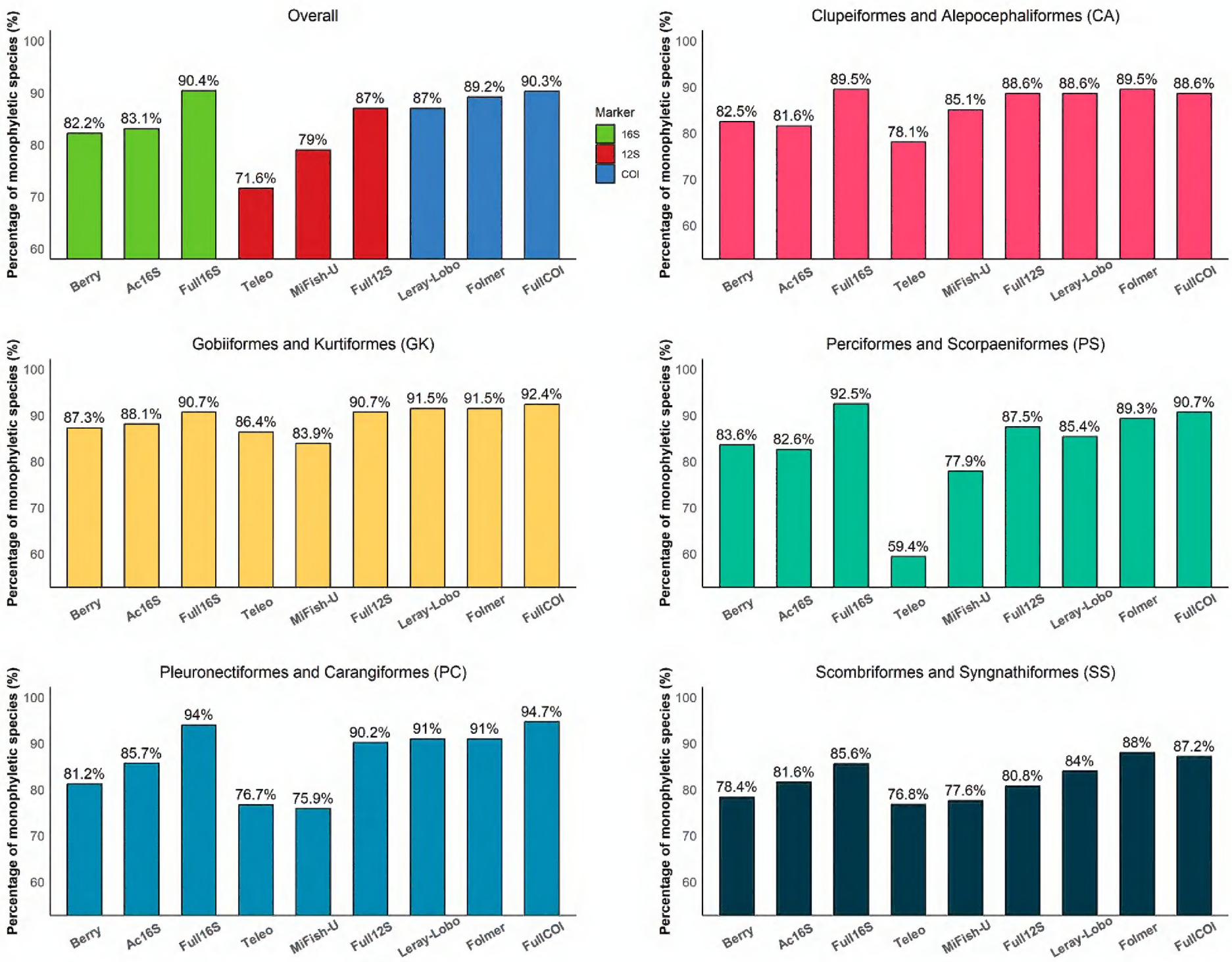

Figure 2. Bar plots displaying the overall and per group percentage of monophyletic species obtained in the analysis using mitochondrial genomes. For each genomic region, CA datasets had 403 sequences and 114 species, GK datasets had 282 sequences and 118 species, PS datasets had 774 sequences and 281 species, PC datasets had 304 sequences and 133 species and SS datasets had 331 sequences and 125 species.

Species discrimination analysis using independent sequences

The overall percentage of monophyletic species obtained for the complete datasets of each genomic region and the proportion of reasons behind non-monophyly, are represented in Table 3. The Leray-Lobo region had the lowest percentage of monophyly (63.5% for 3,314 species and 26,275 sequences), while the full 16S region had the highest (90.7% for 1,069 species and 2,561 sequences). Reasons for non-monophyly, including species sharing haplotypes, intruders, outliers or both intruders and outliers, are also provided in Table 3.

Overall congeneric distance metrics for the datasets comprising independent sequences can be observed in Table 4, along with the number of pairwise comparisons, species, genera and the number of sequences (i.e. duplicate haplotypes within each species were removed). Specific values of average congeneric divergence for each genus analysed using independent sequences are provided in Suppl. material 6.

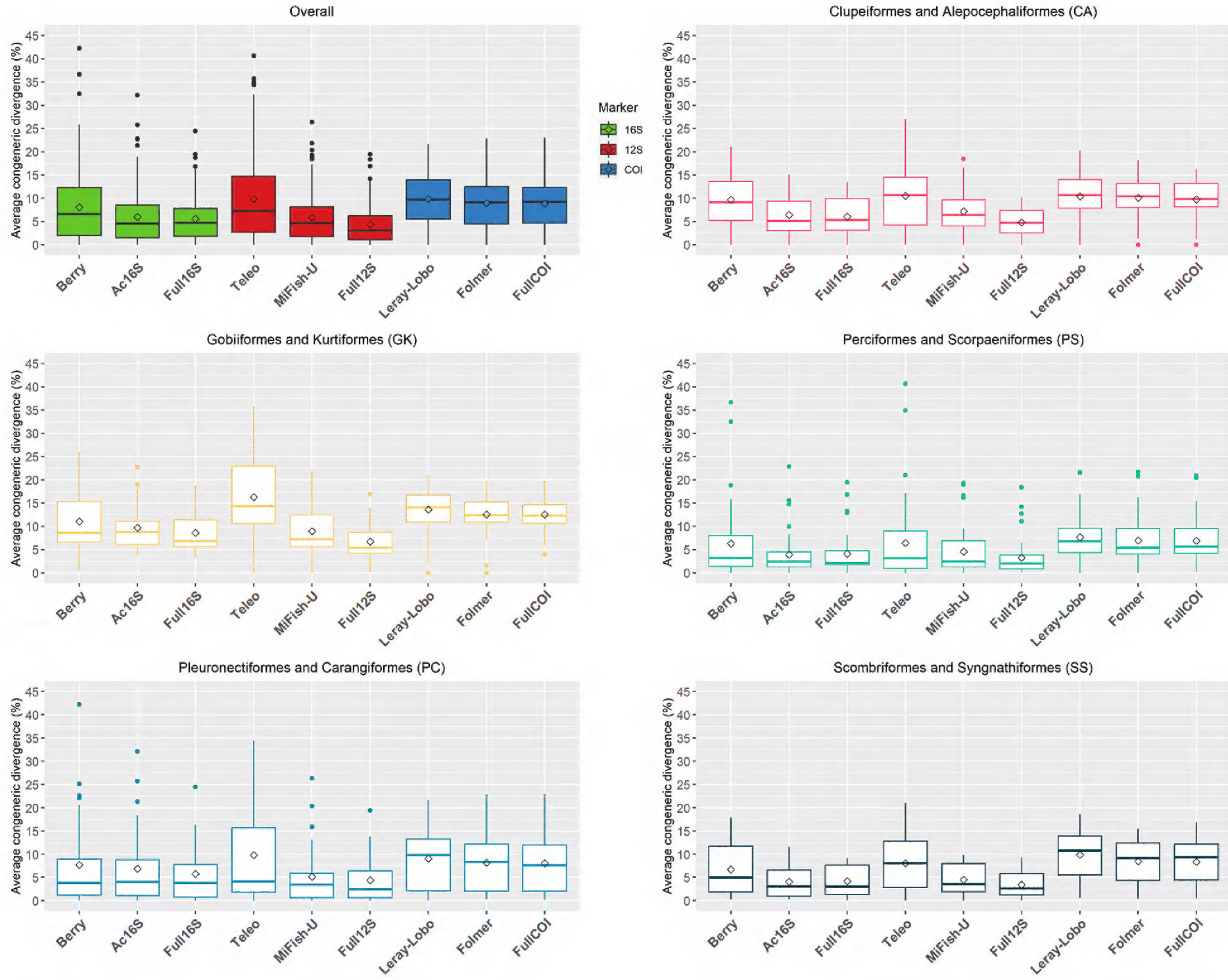

Figure 3. Box plots displaying the overall and per group average congeneric pairwise distance for each genus, calculated in the analysis using mitochondrial genomes. For each genomic region, CA had 344 sequences, 84 species and 23 genera; GK had 178 sequences, 75 species and 27 genera; PS had 564 sequences, 183 species and 43 genera; PC had 204 sequences, 87 species and 30 genera; and SS had 232 sequences, 76 species and 23 genera.

Table 3. Number of sequences, species, average number of sequences per species (SPS), percentage of monophyletic species, percentage of non-monophyletic species that share a haplotype with another species, percentage of non-monophyletic species with at least one intruder sequence, percentage of non-monophyletic species with at least one outlier sequence and percentage of non-monophyletic species with at least one intruder and one outlier sequence.

Region	Sequences	Species	Average SPS	% Monophyletic	% Non-monophyletic sharing haplotype	% Non-monophyletic with intruder	% Non-monophyletic with outlier	% Non-monophyletic with intruder and outlier
Berry	5,752	2,061	2.8	70.7	66.6	39.0	44.8	16.2
Ac16S	3,064	1,521	2.0	83.0	48.3	48.9	42.0	9.1
Full16S	2,561	1,069	2.4	90.7	24.2	54.2	41.0	4.8
Teleo	2,647	1,423	1.9	77.1	70.3	35.1	56.8	8.1
MiFish-U	4,976	2,117	2.4	75.2	67.7	45.9	45.7	8.4
Full12S	2,623	1,104	2.4	86.9	33.8	57.9	35.5	6.6
Leray-Lobo	26,275	3,314	7.9	63.5	49.6	38.6	37.9	23.5
Folmer	34,261	3,310	10.4	64.8	28.2	38.3	38.7	23.0
FullCOI	3,886	1,064	3.7	89.5	16.1	51.0	45.1	3.9

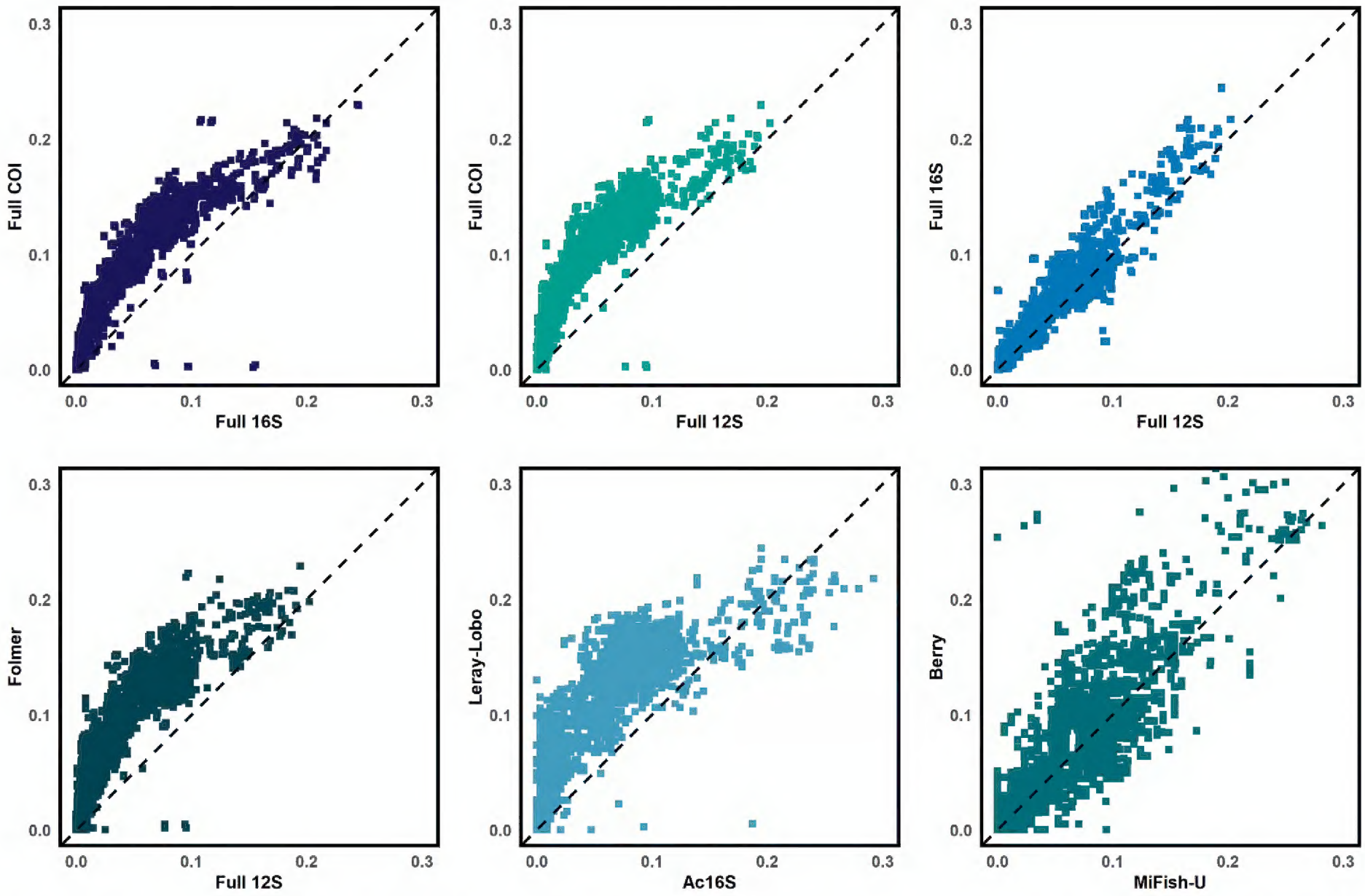

Figure 4. Scatter plots displaying the pairwise congeneric divergence values for six pairs of genomic regions. Each data point represents the pairwise distance between two sequences from different species within the same genus, derived from the same mitochondrial genomes for each pair of regions.

Table 4. Maximum average congeneric divergence (%), mean of the average congeneric divergence (%), standard error (SE), number of pairwise comparisons, number of species, number of genera and number of sequences for all genomic regions analysed.

Region	% Max.	% Mean	SE	Pairwise comparisons	Species	Genera	Sequences
Berry	31.1	8.8	0.4	59,727	1,632	359	3,459
Ac16S	32.1	6.8	0.3	10,486	1,109	287	1,667
Full16S	24.5	6.5	0.4	7,382	725	202	1,332
Teleo	36.9	11.8	0.6	5,797	1,019	252	1,289
MiFish-U	25.1	6.8	0.3	22,704	1,702	366	2,827
Full12S	19.4	5.3	0.3	7,145	759	199	1,359
Leray-Lobo	24.4	11.2	0.3	1,072,424	2,858	495	17,658
Folmer	22.9	10.3	0.3	2,472,090	2,857	497	25,217
FullCOI	23.0	9.5	0.4	206,573	764	200	2,479

Discussion

As long as substantial efforts are made towards the harmonisation of protocols and the completeness and curation of reference libraries, (e)DNA metabarcoding is likely to become one of the main drivers behind biodiversity assessment (Grant et al. 2021). In each step of this methodology's workflow, technical and logistical challenges emerge, which ought to be overcome in the future

(Tsuji et al. 2019). One of the main challenges consists of determining which amplicons should be used to yield the most trustworthy species identifications. In this study, we provide an example of how different amplicons and genetic markers have dissimilar discriminatory power at the species-level, making this an important criterion for selecting the markers to be used in (e)DNA metabarcoding studies. If the main goal in a metabarcoding study is species-level identifications, this may require a multi-marker approach under some scenarios. Moreover, we demonstrate that public reference sequences vary not only in data coverage, but also in their ability to discriminate between species, influenced by potential misidentifications and the inherent complexity of querying sequencing reads against increasingly large datasets.

The COI barcode regions (i.e. Leray-Lobo and Folmer regions) provided the most comprehensive coverage in terms of both the number of species and sequences locally aligned (Table 1), as COI is the standard marker for animal species identification and substantial efforts have been made to complete reference libraries (Weigand et al. 2019; Marques et al. 2021). Nonetheless, the MiFish-U and Berry regions also exhibit substantial coverage for the analysed marine fish taxa, although with lower representation per species (Table 3). Considering that efforts to complete the reference libraries for these regions are bound to continue, we can expect them to become increasingly useful for metabarcoding studies, as the comprehensiveness of their public sequences grows. Moreover, since our study specifically focused on marine fishes, an inclusion of reference sequence data for non-marine species would vastly increase the number of records for all genomic regions, since eDNA metabarcoding is commonly utilised for the study of freshwater and brackish fishes (Shu et al. 2021; García-Machado et al. 2022; Macher et al. 2023).

Species discrimination analysis using mitochondrial genomes

Since our analysis of species discriminatory power was conducted using mitochondrial genomes, we were able to compare it without the biases introduced by potential differences in reference sequence quality between markers. This is because, for each genomic region, we compared sequences obtained from the same individual organisms. Therefore, even if occasional misidentifications might be present, they did not disrupt the discriminatory power comparison between markers and amplicons. At most, misidentifications could potentially diminish the percentage of monophyletic species across all nine genomic regions for a specific group of fishes. In fact, across the nine genomic regions analysed, SS was the group of fishes with the lowest percentage of monophyletic species, while, in contrast, all regions in GK had at least 80% of monophyletic species (Fig. 2). In part, this can be due to the level of accuracy in the morphological identification of some species, with certain taxa being more propitious to misidentifications and ambiguity.

The two longest fragments, full 16S and full COI, were able to discriminate overall the highest proportion of species, in four out of the five groups of fishes. As fragment length increases, sequences are more likely to include a larger number of informative sites that enable species-level discrimination. Similarly, amongst the 12S and 16S amplicons analysed, the longer 16S amplicons had a higher percentage of monophyletic species than the shorter 12S amplicons

in four out of the five groups (GK, PS, PC and SS). While, generally, longer fragments are more capable of discriminating at the species level (Zhang et al. 2020; Polanco et al. 2021), there are exceptions. Despite the full 12S region being the third longest fragment (~ 950 bp), it showed a lower discriminatory power than the Folmer region (~ 658 bp) in all groups. Indeed, the Leray-Lobo region (~ 313 bp) had a higher proportion of monophyletic species than the full 12S region in three out of the five groups (GK, PC and SS). It is also noteworthy that, despite comprising a shorter fragment, the Berry region had a higher percentage of monophyletic species than Ac16S in CA and PS, while MiFish-U (~170 bp) had a higher proportion than the other 12S and 16S amplicons in CA. Even if fragment length is one of the main factors driving species-level discriminatory capability, exceptions are found due to the genetic variability of a particular genomic region, even within the same gene, especially when targeting different taxa. In addition, there were three outliers in the CA and SS datasets, where the Folmer region had a higher proportion of monophyletic species than the full COI region. Since homoplasy is a common phenomenon in mitochondrial DNA due to a high substitution rate, especially for synonymous substitutions (Engstrom et al. 2004; McCracken and Sorenson 2005; Xia 2020), such scenarios may occur.

Higher mean and median values of average congeneric divergence were expected in the case of COI (Fig. 3), where higher substitution rates increase the number of possible informative sites used to discriminate between fish species (Cawthorn et al. 2012; Mohanty et al. 2015). In contrast, high congeneric divergence was not expected of Teleo since it had the lowest overall percentage of monophyletic species. Although this region is highly variable and has a high proportion of pairwise differences, its short length may limit the accumulation of enough informative sites to provide a strong phylogenetic signal. In contrast, the low average congeneric divergence values observed for the full 12S region highlight the noticeable differences in species discriminatory power between markers.

The scatter plots in Fig. 4 further corroborate the tendency of COI regions to exhibit higher congeneric divergence compared to 12S and 16S regions. In contrast, the full 16S and full 12S regions follow a more linear relationship, although the Berry region suggests a higher congeneric divergence than the MiFish-U region. Considering that COI is a protein-coding gene, its higher congeneric divergence relative to the 12S and 16S regions may be partly attributed to intrinsic properties common to protein-coding regions, rather than depending solely on its higher evolutionary rate.

It has been shown that synonymous substitutions are much more prevalent than non-synonymous substitutions for COI in fishes (Ward and Holmes 2007), a pattern that is common amongst many protein-coding genes (Miyata et al 1980). Since purifying selection tends to exert a weaker influence on synonymous substitutions (Graur and Li 1997), this may allow synonymous substitutions to accumulate more freely in closely-related species pairs, leading to a rapid saturation as the evolutionary distance between species increases. This contributes to the non-linear pattern observed when comparing the average congeneric divergence of rRNA genes with protein-coding genes, such as COI. Hence, it is plausible that the higher discriminatory power at the species-level of protein-coding regions, like COI, is partly due to the greater accumulation of synonymous substitutions.

While there is a recognisable difference in discriminatory power between genetic markers and amplicons, other crucial factors should be contemplated for amplicon selection. For instance, technical aspects, such as annealing temperature during PCR or the number of PCR replicates, can influence species detection from environmental samples (Doi et al. 2019). Additionally, despite the high discriminatory power of COI, reported co-amplification of prokaryotes and non-target eukaryotes in fish metabarcoding studies using this marker presents a significant limitation that should be considered, as this issue inevitably leads to a lower percentage of sequencing reads being usable for identifying fish species (Deagle et al. 2014; Collins et al. 2019; Zafeiropoulos et al. 2021). This contrasts with the higher amplification performance and fish specificity reported for 12S and 16S amplicons (Miya et al. 2015; Berry et al. 2017; Polanco et al. 2021; Shu et al. 2021). Although non-target amplification of bacterial DNA has also been reported using the MiFish-U primers (Gold et al. 2021b), the choice of polymerase can mitigate this issue (Kawato et al. 2021).

Overall, Teleo had the lowest percentage of monophyletic species (especially in PS) presumably because it comprises the shortest analysed fragment, with its length averaging below 70 base pairs (Fig. 1). Nonetheless, Teleo can be suitable for specific taxa, as, for example, it discriminated 86.4% of species in the GK dataset (Fig. 2). Since there are technical benefits gained from using shorter fragments (Valentini et al. 2016), for instance, the greater concentration of short fragments in an eDNA sample (Bylemans et al. 2018), this amplicon remains useful for fish detection and identification of species for which its taxonomic resolution proves to be sufficient (Meulenbroek et al. 2022).

Considering the merits and demerits of different markers, certain scenarios may arise where the employment of more than one marker can become necessary. The decision on whether to use a multi-marker approach is ultimately dependent on the nature and objectives of a metabarcoding study, as well as the target ecosystem and sampling locations. For instance, if achieving species-level identifications is not critical for the study's objectives, the usage of just one marker, even if it provides lower discriminatory power, can be effective for ichthyofauna surveys where genus or family-level identifications are sufficient (Thekiso et al. 2023 Açıkbaş et al. 2024). Moreover, if the survey area comprises a smaller body of water, such as a small river or lake, where the number of species is limited or the species can be accounted for *a priori*, a single amplicon with reported merits, such as the MiFish-U region, can be successfully employed for species-level identifications (Fujii et al. 2019; Li et al. 2022). In these scenarios, if the presence of existing close congeneric species is anticipated, it becomes feasible to evaluate the marker, based on its ability to distinguish between these species effectively.

Conversely, in areas with high ichthyofauna diversity, including numerous congeneric species (Stuart-Smith et al. 2013; Granger et al. 2015; Jézéquel et al. 2020), the use of multiple markers is essential for accurate species-level identifications and minimising the risk of both false positives and false negatives (Cordier et al. 2019; Hallam et al. 2021; Wang et al. 2021; Cicala et al. 2024). In such environments, a multi-marker strategy should be adopted as the standard practice, particularly if the goal is to support fisheries management by providing species-level identifications. Studies have indeed shown that multi-marker approaches successfully contribute to species detection in

field surveys, particularly in diverse and complex ecosystems (West et al. 2021; Cicala et al. 2024; Ferreira et al. 2024). However, the adoption of multiple markers is often limited by budgetary considerations, as processing more than one genetic marker incurs higher expenses. Despite the decreasing costs of DNA sequencing in recent years (Dorado et al. 2021; Remec et al. 2021), financial limitations in (e)DNA metabarcoding studies may require a compromise between the number of samples processed and the depth of analysis per sample.

Although utilising several markers is a good starting point when aiming for species-level identifications, there were cases where no marker was able to fully discriminate certain species. For instance, two Clupeiform species, *Alosa alosa* and *Alosa fallax*, were not fully discriminated by any of the genomic regions. These two species are known to hybridise extensively (Alexandrino et al. 2006; Taillebois et al. 2020), hence none of the mitochondrial markers extensively used in fish (e)DNA-based studies is effective in resolving them. Similarly, for the Scombriform genus *Thunnus*, there were several non-monophyletic species across all genomic regions, although this is a group with reported taxonomic ambiguities and occurrence of introgression (Chow et al. 2006; Bayona-Vásquez et al. 2018). Another example within Scombriforms is the genus *Scomber*, where *Scomber japonicus* and *Scomber colias* were not discriminated by any of the genomic regions. Nevertheless, according to an analysis of available COI barcode fragment data (Trucco and Buratti 2017), these two species can be discriminated although displaying comparatively low genetic Kimura 2-Parameter (K2P; Kimura (1980)) distances (0.019). The recognition of the two species was only recently proposed (Scoles et al. 1998) and their geographic distributions clarified, with the former occurring in the Pacific and the latter in the Eastern Atlantic. This may lead to mislabelling of sequence records in databases, that did not assume the updated taxonomy, even in the case of full mitochondrial genome data. The occurrence of inaccuracies in the genetic data deposited in public databases has been increasingly recognised as a peril for the accuracy of (e)DNA-based monitoring and addressed in various studies (Arranz et al. 2020; Pentinsaari et al. 2020; Fontes et al. 2021; Radulovici et al. 2021; Claver et al. 2023).

Species discrimination analysis using independent sequences

For the analysis using each genomic region's independent sequences, Table 3 shows that the percentages of monophyletic species greatly differ from the initial analysis using mitogenomes (Fig. 2). The complete datasets vary considerably in the number of sequences and species and especially in the number of sequences per species (Table 3); therefore, they are not comparable in terms of discriminatory power of the genomic region itself. Similarly, although the average congeneric divergence values for the complete datasets (Table 4) are close to those found in the mitogenomes analysis (Fig. 3), they are not directly comparable across regions. Regardless, since GenBank is often used as the direct source of reference sequence data in eDNA metabarcoding (Jeunen et al. 2019; Kirse et al. 2021; Lamy et al. 2021; Sato et al. 2021), the complete dataset analysis provides insights into the reliability of these public datasets for taxonomic assignment.

It is evident that, as the number of sequences and species increases, there is a reduction in the proportion of monophyletic species, even if the marker itself

is usually capable of discriminating at the species-level. Table 3 shows that the Leray-Lobo and Folmer region had the lowest percentage of monophyletic species, in opposition to the analysis with mitogenomes, where they were amongst the genomic regions with the highest discriminatory power (Fig. 2). However, they include a higher number of sequences, species and sequences per species (Table 3) than any other fragment and they also have a higher percentage of non-monophyletic species with both intruders and outliers. The presence of both intruders and outliers can be indicative of more numerous misidentifications, which may be more frequent in genomic regions for which there is a higher number of sequence data. For example, in the Leray-Lobo dataset, *Epinephelus areolatus* had 79 sequences that would be monophyletic if not for a single outlier sequence (likely a misidentification) and one intruder species, *Epinephelus chlorostigma*, which had 10 total sequences in the tree. Conversely, in the Teleo dataset, both species were monophyletic, although they comprise only three and two sequences, respectively. Even though all genomic regions will be impacted by misidentifications, regions with higher number of sequences per species will inevitably accumulate more instances.

The datasets of shorter fragments such as Berry, Teleo and MiFish-U have a higher percentage of non-monophyletic species sharing a haplotype with another species (66.6%, 70.3% and 67.7%, respectively), further corroborating their lower discriminatory power. During taxonomic assignment in (e)DNA metabarcoding, one ambiguous sequence can be enough to obtain a false positive or false negative. Therefore, if a metabarcoding-based ichthyofauna survey targets closely-related congeneric species, the standard approach should be to use several fragments to allow the downstream rectification of potential ambiguities. In addition to the species discriminatory power, it is essential to consider the importance of auditing and curating reference sequence databases to obtain accurate taxonomic assignments. Although the reference libraries for COI have been subject to many auditing and curation efforts (Fontes et al. 2021; Porter and Hajibabaei 2021; Radulovici et al. 2021; Claver et al. 2023; Lavrador et al. 2023; Meglécz 2023), similar efforts should be extended to other markers used in (e)DNA (meta)barcoding. Efforts have indeed begun for these markers (Arranz et al. 2020; Claver et al. 2023), but the increasing volume of reference sequence data emphasises the need for continued and expanded curation efforts.

Conclusions

Our comprehensive analysis of the species discrimination ability of the most common markers employed in marine fish (e)DNA metabarcoding helped benchmark the limitations and differences amongst them. Considering how different amplicons and markers yield different results both in terms of reference database completeness and discriminatory power, one should be careful not to overestimate the current effectiveness of (e)DNA metabarcoding, especially when using only one amplicon or marker under scenarios of high ichthyofauna diversity. While gaps in reference sequence data coverage between different fragments can be mitigated in the future, differences in species discriminatory power have inherent and immutable biological reasons and may lead to systematic loss of species-level data in (e)DNA-based studies. Hence, we warn of the relevance of considering species discrimination

ability as one of the main criteria when selecting amplicons for marine fish metabarcoding studies.

Even though COI remains the standard genetic marker for animal species identification, with its regions having the highest discriminatory power, it is important to recognise the advantages and disadvantages of each amplicon since no genomic region is universally optimal. Thus, careful consideration is fundamental, as each region has its own set of limitations and affinities and no single marker is perfectly suited to all studies or ecosystems. The choice of amplicon must be tailored to the specific requirements of the research question and the ecological context.

Therefore, to maximise species identification accuracy and minimise the risk of false positives and negatives, we advocate the use of multiple target regions in (e)DNA metabarcoding studies aiming for species-level identifications of closely-related marine fish taxa, while carefully integrating and curating the information obtained. A multi-marker approach will help resolve the ambiguities that may arise from certain genomic regions not having enough discriminatory power, while compensating for the technical limitations of certain amplicons. Especially for the potential use in fishery stock assessment, where precise discrimination amongst closely-related species is of extreme importance, we recommend the use of multiple amplicons as the standard approach to harness the full potential of (e)DNA metabarcoding.

Additional information

Conflict of interest

The authors have declared that no competing interests exist.

Ethical statement

No ethical statement was reported.

Funding

This work was funded by the project "A-FISH-DNA-SCAN: Cutting-edge DNA-based approaches for improved monitoring and management of fisheries resources along Magellan-Elcano's Atlantic route" (CIRCNA/BRB/0156/2019; http://doi.org/10.54499/CIRCNA/BRB/0156/2019) funded by the Portuguese Foundation of Science and Technology (FCT, I.P.), by the CBMA "Contrato-Programa" (UIDB/04050/2020; https://doi.org/10.54499/UIDB/04050/2020) and by the ARNET "Contrato-Programa"(LA/P/0069/2020; https://doi.org/10.54499/LA/P/0069/2020), funded by national funds through FCT I.P. J.T.F. (UI/BD/150910/2021; https://doi.org/10.54499/UI/BD/150910/2021) is supported by the Collaboration Protocol for Financing the Multiannual Research Grants Plan for Doctoral Students with financial support from FCT I.P. and the European Social Fund under the Northern Regional Operational Program – Norte2020. K.K. is supported by JSPS KAKENHI, Grant Number JP20K06767.

Author contributions

Conceptualisation, J.T.F., P.S. and F.O.C.; methodology, J.T.F, K.K., P.S. and F.O.C; software, J.T.F., K.K. and R.P., validation, J.T.F, K.K. and R.P., formal analysis, J.T.F, writing – original draft, J.T.F, P.S. and F.O.C, writing – review and editing, J.T.F, K.K., R.P., P.S. and F.O.C.

Author ORCIDs

João T. Fontes https://orcid.org/0000-0002-8766-4779
Kazutaka Katoh https://orcid.org/0000-0003-4133-8393
Pedro Soares https://orcid.org/0000-0002-2807-690X
Filipe O. Costa https://orcid.org/0000-0001-5398-3942

Data availability

All of the data that support the findings of this study are available in the main text or Supplementary Information.

References

Açıkbaş AHO, Narisoko H, Huerlimann R, Nishitsuji K, Satoh N, Reimer JD, Ravasi T (2024) Fish and coral assemblages of a highly isolated oceanic island: The first eDNA survey of the Ogasawara Islands. Environmental DNA 6(1): e509. https://doi.org/10.1002/edn3.509

Alexandrino P, Faria R, Linhares D, Castro F, Le Corre M, Sabatié R, Baglinière JL, Weiss S (2006) Interspecific differentiation and intraspecific substructure in two closely related clupeids with extensive hybridization, *Alosa alosa* and *Alosa fallax*. Journal of Fish Biology 69(sb): 242–259. https://doi.org/10.1111/j.1095-8649.2006.01289.x

Arranz V, Pearman WS, Aguirre JD, Liggins L (2020) MARES, a replicable pipeline and curated reference database for marine eukaryote metabarcoding. Scientific Data 7(1): 209. https://doi.org/10.1038/s41597-020-0549-9

Bayona-Vásquez NJ, Glenn TC, Uribe-Alcocer M, Pecoraro C, Díaz-Jaimes P (2018) Complete mitochondrial genome of the yellowfin tuna (*Thunnus albacares*) and the blackfin tuna (*Thunnus atlanticus*): Notes on mtDNA introgression and paraphyly on tunas. Conservation Genetics Resources 10(4): 697–699. https://doi.org/10.1007/s12686-017-0904-0

Berry TE, Osterrieder SK, Murray DC, Coghlan ML, Richardson AJ, Grealy AK, Stat M, Bejder L, Bunce M (2017) DNA metabarcoding for diet analysis and biodiversity: A case study using the endangered Australian sea lion (*Neophoca cinerea*). Ecology and Evolution 7(14): 5435–5453. https://doi.org/10.1002/ece3.3123

Bodenhofer U, Bonatesta E, Horejš-Kainrath C, Hochreiter S (2015) msa: An R package for multiple sequence alignment. Bioinformatics (Oxford, England) 31(24): 3997–3999. https://doi.org/10.1093/bioinformatics/btv494

Buhay JE (2009) "COI-like" sequences are becoming problematic in molecular systematic and DNA barcoding studies. Journal of Crustacean Biology 29(1): 96–110. https://doi.org/10.1651/08-3020.1

Bylemans J, Furlan EM, Gleeson DM, Hardy CM, Duncan RP (2018) Does size matter? An experimental evaluation of the relative abundance and decay rates of aquatic environmental DNA. Environmental Science & Technology 52(11): 6408–6416. https://doi.org/10.1021/acs.est.8b01071

Cawthorn D-M, Steinman HA, Witthuhn RC (2012) Evaluation of the 16S and 12S rRNA genes as universal markers for the identification of commercial fish species in South Africa. Gene 491(1): 40–48. https://doi.org/10.1016/j.gene.2011.09.009

Chow S, Nakagawa T, Suzuki N, Takeyama H, Matsunaga T (2006) Phylogenetic relationships among *Thunnus* species inferred from rDNA ITS1 sequence. Journal of Fish Biology 68(A): 24–35. https://doi.org/10.1111/j.0022-1112.2006.00945.x

Cicala D, Maiello G, Fiorentino F, Garofalo G, Massi D, Sbrana A, Mariani S, D'Alessandro S, Stefani M, Perrodin L, Russo T (2024) Spatial analysis of demersal food webs through integration of eDNA metabarcoding with fishing activities. Frontiers in Marine Science 10: 1209093. https://doi.org/10.3389/fmars.2023.1209093

Claver C, Canals O, de Amézaga LG, Mendibil I, Rodriguez-Ezpeleta N (2023) An automated workflow to assess completeness and curate GenBank for environmental DNA metabarcoding: The marine fish assemblage as case study. Environmental DNA 5(4): 634–647. https://doi.org/10.1002/edn3.433

Collins RA, Bakker J, Wangensteen OS, Soto AZ, Corrigan L, Sims DW, Genner MJ, Mariani S (2019) Non-specific amplification compromises environmental DNA metabarcoding with COI. Methods in Ecology and Evolution 10(11): 1985–2001. https://doi.org/10.1111/2041-210X.13276

Collins RA, Trauzzi G, Maltby KM, Gibson TI, Ratcliffe FC, Hallam J, Rainbird S, Maclaine J, Henderson PA, Sims DW, Mariani S, Genner MJ (2021) Meta-Fish-Lib: A generalised, dynamic DNA reference library pipeline for metabarcoding of fishes. Journal of Fish Biology 99(4): 1446–1454. https://doi.org/10.1111/jfb.14852

Cordier T, Frontalini F, Cermakova K, Apothéloz-Perret-Gentil L, Treglia M, Scantamburlo E, Bonamin V, Pawlowski J (2019) Multi-marker eDNA metabarcoding survey to assess the environmental impact of three offshore gas platforms in the North Adriatic Sea (Italy). Marine Environmental Research 146: 24–34. https://doi.org/10.1016/j.marenvres.2018.12.009

De Mandal S, Chhakchhuak L, Gurusubramanian G, Kumar NS (2014) Mitochondrial markers for identification and phylogenetic studies in insects – A Review. DNA Barcodes 2(1): 1–9. https://doi.org/10.2478/dna-2014-0001

Deagle BE, Jarman SN, Coissac E, Pompanon F, Taberlet P (2014) DNA metabarcoding and the cytochrome c oxidase subunit I marker: Not a perfect match. Biology Letters 10(9): 20140562. https://doi.org/10.1098/rsbl.2014.0562

Deiner K, Bik HM, Mächler E, Seymour M, Lacoursière-Roussel A, Altermatt F, Creer S, Bista I, Lodge DM, De Vere N, Pfrender ME, Bernatchez L (2017) Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. Molecular Ecology 26(21): 5872–5895. https://doi.org/10.1111/mec.14350

Doi H, Fukaya K, Oka SI, Sato K, Kondoh M, Miya M (2019) Evaluation of detection probabilities at the water-filtering and initial PCR steps in environmental DNA metabarcoding using a multispecies site occupancy model. Scientific Reports 9(1): 3581. https://doi.org/10.1038/s41598-019-40233-1

Dorado G, Gálvez S, Rosales TE, Vásquez VF, Hernández P (2021) Analyzing modern biomolecules: The revolution of nucleic-acid sequencing – Review. Biomolecules 11(8): 1111. https://doi.org/10.3390/biom11081111

Engstrom TN, Shaffer HB, McCord WP (2004) Multiple data sets, high homoplasy, and the phylogeny of softshell turtles (Testudines: Trionychidae). Systematic Biology 53(5): 693–710. https://doi.org/10.1080/10635150490503053

Evans NT, Olds BP, Renshaw MA, Turner CR, Li Y, Jerde CL, Mahon AR, Pfrender ME, Lamberti GA, Lodge DM (2016) Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. Molecular Ecology Resources 16(1): 29–41. https://doi.org/10.1111/1755-0998.12433

Ferreira AO, Azevedo OM, Barroso C, Duarte S, Egas C, Fontes JT, Ré P, Piecho-Santos AM, Costa FO (2024) Multi-marker DNA metabarcoding for precise species identification in ichthyoplankton samples. Scientific Reports 14(1): 19772. https://doi.org/10.1038/s41598-024-69963-7

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology 3: 294–299.

Fontes JT, Vieira PE, Ekrem T, Soares P, Costa FO (2021) BAGS: An automated Barcode, Audit & Grade System for DNA barcode reference libraries. Molecular Ecology Resources 21(2): 573–583. https://doi.org/10.1111/1755-0998.13262

Fujii K, Doi H, Matsuoka S, Nagano M, Sato H, Yamanaka H (2019) Environmental DNA metabarcoding for fish community analysis in backwater lakes: A comparison of capture methods. PLoS One 14(1): e0210357. https://doi.org/10.1371/journal.pone.0210357

García-Machado E, Laporte M, Normandeau E, Hernández C, Côté G, Paradis Y, Mingelbier M, Bernatchez L (2022) Fish community shifts along a strong fluvial environmental gradient revealed by eDNA metabarcoding. Environmental DNA 4(1): 117–134. https://doi.org/10.1002/edn3.221

Gold Z, Sprague J, Kushner DJ, Zerecero Marin E, Barber PH (2021a) eDNA metabarcoding as a biomonitoring tool for marine protected areas. PLoS One 16(2): e0238557. https://doi.org/10.1371/journal.pone.0238557

Gold Z, Curd EE, Goodwin KD, Choi ES, Frable BW, Thompson AR, Walker Jr HJ, Burton RS, Kacev D, Martz LD, Barber PH (2021b) Improving metabarcoding taxonomic assignment: A case study of fishes in a large marine ecosystem. Molecular Ecology Resources 21(7): 2546–2564. https://doi.org/10.1111/1755-0998.13450

Granger V, Fromentin J-M, Bez N, Relini G, Meynard CN, Gaertner J-C, Maiorano P, Ruiz CG, Follesa C, Gristina M, Peristeraki P (2015) Large-scale spatio-temporal monitoring highlights hotspots of demersal fish diversity in the Mediterranean Sea. Progress in Oceanography 130: 65–74. https://doi.org/10.1016/j.pocean.2014.10.002

Grant DM, Brodnicke OB, Evankow AM, Ferreira AO, Fontes JT, Hansen AK, Jensen MR, Kalaycı TE, Leeper A, Patil SK, Prati S, Reunamo A, Roberts AJ, Shigdel R, Tyukosova V, Bendiksby M, Blaalid R, Costa FO, Hollingsworth PM, Stur E, Ekrem T (2021) The future of DNA barcoding: Reflections from early career researchers. Diversity 13(7): 313. https://doi.org/10.3390/d13070313

Graur D, Li W-H (1997) Molecular evolution. Sinauer Associates, Sunderland, MA.

Hajibabaei M, Porter TM, Wright M, Rudar J (2019) COI metabarcoding primer choice affects richness and recovery of indicator taxa in freshwater systems. PLoS One 14(9): e0220953. https://doi.org/10.1371/journal.pone.0220953

Hallam J, Clare EL, Jones JI, Day JJ (2021) Biodiversity assessment across a dynamic riverine system: A comparison of eDNA metabarcoding versus traditional fish surveying methods. Environmental DNA 3(6): 1247–1266. https://doi.org/10.1002/edn3.241

Hansen BK, Bekkevold D, Clausen LW, Nielsen EE (2018) The sceptical optimist: Challenges and perspectives for the application of environmental DNA in marine fisheries. Fish and Fisheries 19(5): 751–768. https://doi.org/10.1111/faf.12286

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. Proceedings. Biological Sciences 270(1512): 313–321. https://doi.org/10.1098/rspb.2002.2218

Holstein J (2018) worms: Retriving Aphia Information from World Register of Marine Species. R package version 0.2.2. https://cran.r-project.org/package=worms

Jeunen G-J, Knapp M, Spencer HG, Lamare MD, Taylor HR, Stat M, Bunce M, Gemmell NJ (2019) Environmental DNA (eDNA) metabarcoding reveals strong discrimination among diverse marine habitats connected by water movement. Molecular Ecology Resources 19(2): 426–438. https://doi.org/10.1111/1755-0998.12982

Jézéquel C, Tedesco PA, Darwall W, Dias MS, Frederico RG, Hidalgo M, Hugueny B, Maldonado-Ocampo J, Martens K, Ortega H, Torrente-Vilara G, Zuanon J, Oberdorff T (2020) Freshwater fish diversity hotspots for conservation priorities in the Amazon Basin. Conservation Biology 34(4): 956–965. https://doi.org/10.1111/cobi.13466

Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL (2008) NCBI BLAST: A better web interface. Nucleic Acids Research 36(Web Server): W5–W9. https://doi.org/10.1093/nar/gkn201

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30(14): 3059–3066. https://doi.org/10.1093/nar/gkf436

Kawato M, Yoshida T, Miya M, Tsuchida S, Nagano Y, Nomura M, Yabuki A, Fujiwara Y, Fujikura K (2021) Optimization of environmental DNA extraction and amplification methods for metabarcoding of deep-sea fish. MethodsX 8: 101238. https://doi.org/10.1016/j.mex.2021.101238

Kiełbasa SM, Wan R, Sato K, Horton P, Frith MC (2011) Adaptive seeds tame genomic sequence comparison. Genome Research 21(3): 487–493. https://doi.org/10.1101/gr.113985.110

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16(2): 111–120. https://doi.org/10.1007/BF01731581

Kirse A, Bourlat SJ, Langen K, Fonseca VG (2021) Unearthing the potential of soil eDNA metabarcoding – Towards best practice advice for invertebrate biodiversity assessment. Frontiers in Ecology and Evolution 9: 337. https://doi.org/10.3389/fevo.2021.630560

Lamy T, Pitz KJ, Chavez FP, Yorke CE, Miller RJ (2021) Environmental DNA reveals the fine-grained and hierarchical spatial structure of kelp forest fish communities. Scientific Reports 11(1): 14439. https://doi.org/10.1038/s41598-021-93859-5

Landi M, Dimech M, Arculeo M, Biondo G, Martins R, Carneiro M, Carvalho GR, Lo Brutto S, Costa FO (2014) DNA barcoding for species assignment: The case of Mediterranean marine fishes. PLoS One 9(9): e106135. https://doi.org/10.1371/journal.pone.0106135

Lavrador AS, Fontes JT, Vieira PE, Costa FO, Duarte S (2023) Compilation, revision, and annotation of DNA barcodes of marine invertebrate non-indigenous species (NIS) occurring in European coastal regions. Diversity 15(2): 174. https://doi.org/10.3390/d15020174

Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N, Ranwez V, Boehm JT, Machida RJ (2013) A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: Application for characterizing coral reef fish gut contents. Frontiers in Zoology 10(1): 1–14. https://doi.org/10.1186/1742-9994-10-34

Li H, Yang F, Zhang R, Liu S, Yang Z, Lin L, Ye S (2022) Environmental DNA metabarcoding of fish communities in a small hydropower dam reservoir: A comparison between the eDNA approach and established fishing methods. Journal of Freshwater Ecology 37(1): 341–362. https://doi.org/10.1080/02705060.2022.2086181

Liu M, Clarke LJ, Baker SC, Jordan GJ, Burridge CP (2020) A practical guide to DNA metabarcoding for entomological ecologists. Ecological Entomology 45(3): 373–385. https://doi.org/10.1111/een.12831

Liu B, Yang J-W, Liu B-S, Zhang N, Guo L, Guo H-Y, Zhang D-C (2022) Detection and identification of marine fish mislabeling in Guangzhou's supermarkets and sushi restaurants using DNA barcoding. Journal of Food Science 87(6): 2440–2449. https://doi.org/10.1111/1750-3841.16150

Lobo J, Costa PM, Teixeira MA, Ferreira MS, Costa MH, Costa FO (2013) Enhanced primers for amplification of DNA barcodes from a broad range of marine metazoans. BMC Ecology 13(1): 34. https://doi.org/10.1186/1472-6785-13-34

Macher T-H, Schütz R, Yildiz A, Beermann AJ, Leese F (2023) Evaluating five primer pairs for environmental DNA metabarcoding of Central European fish species based on mock communities. Metabarcoding and Metagenomics 7: e103856. https://doi.org/10.3897/mbmg.7.103856

Machida RJ, Leray M, Ho S-L, Knowlton N (2017) Metazoan mitochondrial gene sequence reference datasets for taxonomic assignment of environmental samples. Scientific Data 4(1): 1–7. https://doi.org/10.1038/sdata.2017.27

Marques V, Milhau T, Albouy C, Dejean T, Manel S, Mouillot D, Juhel J (2021) GAPeDNA: Assessing and mapping global species gaps in genetic databases for eDNA metabarcoding. Diversity & Distributions 27(10): 1880–1892. https://doi.org/10.1111/ddi.13142

Mathon L, Valentini A, Guérin P-E, Normandeau E, Noel C, Lionnet C, Boulanger E, Thuiller W, Bernatchez L, Mouillot D, Dejean T, Manel S (2021) Benchmarking bioinformatic tools for fast and accurate eDNA metabarcoding species identification. Molecular Ecology Resources 21(7): 2565–2579. https://doi.org/10.1111/1755-0998.13430

McCracken KG, Sorenson MD (2005) Is homoplasy or lineage sorting the source of incongruent mtDNA and nuclear gene trees in the stiff-tailed ducks (Nomonyx-Oxyura)? Systematic Biology 54(1): 35–55. https://doi.org/10.1080/10635150590910249

Medina M, Walsh PJ (2000) Molecular systematics of the order Anaspidea based on mitochondrial DNA sequence (12S, 16S, and COI). Molecular Phylogenetics and Evolution 15(1): 41–58. https://doi.org/10.1006/mpev.1999.0736

Meglécz E (2023) COInr and mkCOInr: Building and customizing a nonredundant barcoding reference database from BOLD and NCBI using a semi-automated pipeline. Molecular Ecology Resources 23(4): 933–945. https://doi.org/10.1111/1755-0998.13756

Meulenbroek P, Hein T, Friedrich T, Valentini A, Erős T, Schabuss M, Zornig H, Lenhardt M, Pekarik L, Jean P, Dejean T, Pont D (2022) Sturgeons in large rivers: Detecting the near-extinct needles in a haystack via eDNA metabarcoding from water samples. Biodiversity and Conservation 31(11): 2817–2832. https://doi.org/10.1007/s10531-022-02459-w

Miya M, Sato Y, Fukunaga T, Sado T, Poulsen JY, Sato K, Minamoto T, Yamamoto S, Yamanaka H, Araki H, Kondoh M, Iwasaki W (2015) MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: Detection of more than 230 subtropical marine species. Royal Society Open Science 2(7): 150088. https://doi.org/10.1098/rsos.150088

Miyata T, Yasunaga T, Nishida T (1980) Nucleotide sequence divergence and functional constraint in mRNA evolution. Proceedings of the National Academy of Sciences of the United States of America 77(12): 7328–7332. https://doi.org/10.1073/pnas.77.12.7328

Mohanty M, Jayasankar P, Sahoo L, Das P (2015) A comparative study of COI and 16 S rRNA genes for DNA barcoding of cultivable carps in India. Mitochondrial DNA 26(1): 79–87. https://doi.org/10.3109/19401736.2013.823172

Oliveira L, Knebelsberger T, Landi M, Soares P, Raupach M, Costa FO (2016) Assembling and auditing a comprehensive DNA barcode reference library for European marine fishes. Journal of Fish Biology 89(6): 2741–2754. https://doi.org/10.1111/jfb.13169

Othman N, Haris H, Fatin Z, Najmuddin MF, Sariyati NH, Md-Zain BM, Abdul-Latiff MAB (2021) A review on environmental DNA (eDNA) metabarcoding markers for wildlife monitoring research. IOP Conference Series: Earth and Environmental Science. IOP Publishing 736: 012054. https://doi.org/10.1088/1755-1315/736/1/012054

Paradis E, Schliep K (2019) ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics (Oxford, England) 35(3): 526–528. https://doi.org/10.1093/bioinformatics/bty633

Pentinsaari M, Ratnasingham S, Miller SE, Hebert PDN (2020) BOLD and GenBank revisited – Do identification errors arise in the lab or in the sequence libraries? PLoS One 15(4): e0231814. https://doi.org/10.1371/journal.pone.0231814

Polanco FA, Richards E, Flück B, Valentini A, Altermatt F, Brosse S, Walser J-C, Eme D, Marques V, Manel S (2021) Comparing the performance of 12S mitochondrial primers for fish environmental DNA across ecosystems. Environmental DNA 3(6): 1113–1127. https://doi.org/10.1002/edn3.232

Porter TM, Hajibabaei M (2018) Over 2.5 million COI sequences in GenBank and growing. PLoS One 13(9): e0200177. https://doi.org/10.1371/journal.pone.0200177

Porter TM, Hajibabaei M (2021) Profile hidden Markov model sequence analysis can help remove putative pseudogenes from DNA barcoding and metabarcoding datasets. BMC Bioinformatics 22(1): 256. https://doi.org/10.1186/s12859-021-04180-x

R Core Team (2023) R: a language and environment for statistical computing. R Foundation for Statistical Computing. https://www.r-project.org/

Rabosky DL, Chang J, Title PO, Cowman PF, Sallan L, Friedman M, Kaschner K, Garilao C, Near TJ, Coll M, Alfaro ME (2018) An inverse latitudinal gradient in speciation rate for marine fishes. Nature 559(7714): 392–395. https://doi.org/10.1038/s41586-018-0273-1

Radulovici AE, Vieira PE, Duarte S, Teixeira MAL, Borges LMS, Deagle BE, Majaneva S, Redmond N, Schultz JA, Costa FO (2021) Revision and annotation of DNA barcode records for marine invertebrates: Report of the 8th iBOL conference hackathon. Metabarcoding and Metagenomics 5: e67862. https://doi.org/10.3897/mbmg.5.67862

Remec ZI, Trebusak Podkrajsek K, Repic Lampret B, Kovac J, Groselj U, Tesovnik T, Battelino T, Debeljak M (2021) Next-generation sequencing in newborn screening: A review of current state. Frontiers in Genetics 12: 662254. https://doi.org/10.3389/fgene.2021.662254

Sato M, Inoue N, Nambu R, Furuichi N, Imaizumi T, Ushio M (2021) Quantitative assessment of multiple fish species around artificial reefs combining environmental DNA metabarcoding and acoustic survey. Scientific Reports 11(1): 19477. https://doi.org/10.1038/s41598-021-98926-5

Sayers EW, Cavanaugh M, Clark K, Pruitt KD, Schoch CL, Sherry ST, Karsch-Mizrachi I (2021) GenBank. Nucleic Acids Research 49(D1): D92–D96. https://doi.org/10.1093/nar/gkaa1023

Schoch CL, Ciufo S, Domrachev M, Hotton CL, Kannan S, Khovanskaya R, Leipe D, Mcveigh R, O'Neill K, Robbertse B, Sharma S, Soussov V, Sullivan JP, Sun L, Turner S, Karsch-Mizrachi I (2020) NCBI Taxonomy: A comprehensive update on curation, resources and tools. Database (Oxford) 2020: baaa062. https://doi.org/10.1093/database/baaa062

Schwery O, O'Meara BC (2016) MonoPhy: A simple R package to find and visualize monophyly issues. PeerJ. Computer Science 2: e56. https://doi.org/10.7717/peerj-cs.56

Scoles DS, Collette BB, Graves JE (1998) Global phylogeography of mackerels of the genus Scomber. Fish Bulletin.

Shu L, Ludwig A, Peng Z (2021) Environmental DNA metabarcoding primers for freshwater fish detection and quantification: In silico and in tanks. Ecology and Evolution 11(12): 8281–8294. https://doi.org/10.1002/ece3.7658

Stoeckle MY, Adolf J, Charlop-Powers Z, Dunton KJ, Hinks G, VanMorter SM (2021) Trawl and eDNA assessment of marine fish diversity, seasonality, and relative abundance

in coastal New Jersey, USA. ICES Journal of Marine Science 78(1): 293–304. https://doi.org/10.1093/icesjms/fsaa225

Stuart-Smith RD, Bates AE, Lefcheck JS, Duffy JE, Baker SC, Thomson RJ, Stuart-Smith JF, Hill NA, Kininmonth SJ, Airoldi L, Becerro MA, Campbell SJ, Dawson TP, Navarrete SA, Soler GA, Strain EMA, Willis TJ, Edgar GJ (2013) Integrating abundance and functional traits reveals new global hotspots of fish diversity. Nature 501(7468): 539–542. https://doi.org/10.1038/nature12529

Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. Molecular Ecology 21(8): 2045–2050. https://doi.org/10.1111/j.1365-294X.2012.05470.x

Taillebois L, Sabatino S, Manicki A, Daverat F, Nachón DJ, Lepais O (2020) Variable outcomes of hybridization between declining *Alosa alosa* and *Alosa fallax*. Evolutionary Applications 13(4): 636–651. https://doi.org/10.1111/eva.12889

Tamura K, Stecher G, Kumar S (2021) MEGA11: Molecular evolutionary genetics analysis version 11. Molecular Biology and Evolution 38(7): 3022–3027. https://doi.org/10.1093/molbev/msab120

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research 22(22): 4673–4680. https://doi.org/10.1093/nar/22.22.4673

Trucco MI, Buratti CC (2017) Taxonomic review of Argentine mackerel *Scomber japonicus* (Houttuyn, 1782) by phylogenetic analysis. Molecular Biology Research Communications 6: 141. https://doi.org/10.22099/mbrc.2017.25981.1276

Tsuji S, Takahara T, Doi H, Shibata N, Yamanaka H (2019) The detection of aquatic macroorganisms using environmental DNA analysis – A review of methods for collection, extraction, and detection. Environmental DNA 1(2): 99–108. https://doi.org/10.1002/edn3.21

Valentini A, Taberlet P, Miaud C, Civade R, Herder J, Thomsen PF, Bellemain E, Besnard A, Coissac E, Boyer F, Gaboriaud C, Jean P, Poulet N, Roset N, Copp GH, Geniez P, Pont D, Argillier C, Baudoin J, Peroux T, Crivelli AJ, Olivier A, Acqueberge M, Le Brun M, Møller PR, Willerslev E, Dejean T (2016) Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. Molecular Ecology 25(4): 929–942. https://doi.org/10.1111/mec.13428

Wang S, Yan Z, Hänfling B, Zheng X, Wang P, Fan J, Li J (2021) Methodology of fish eDNA and its applications in ecology and environment. The Science of the Total Environment 755: 142622. https://doi.org/10.1016/j.scitotenv.2020.142622

Ward RD, Holmes BH (2007) An analysis of nucleotide and amino acid variability in the barcode region of cytochrome c oxidase I (cox1) in fishes. Molecular Ecology Notes 7(6): 899–907. https://doi.org/10.1111/j.1471-8286.2007.01886.x

Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding Australia's fish species. Philosophical Transactions of the Royal Society B: Biological Sciences 360: 1847–1857. https://doi.org/10.1111/j.1471-8286.2007.01886.x

Weigand H, Beermann AJ, Čiampor F, Costa FO, Csabai Z, Duarte S, Geiger MF, Grabowski M, Rimet F, Rulik B, Strand M, Szucsich N, Weigand AM, Willassen E, Wyler SA, Bouchez A, Borja A, Čiamporová-Zaťovičová Z, Ferreira S, Dijkstra KB, Eisendle U, Freyhof J, Gadawski P, Graf W, Haegerbaeumer A, van der Hoorn BB, Japoshvili B, Keresztes L, Keskin E, Leese F, Macher JN, Mamos T, Paz G, Pešić V, Pfannkuchen DM, Pfannkuchen MA, Price BW, Rinkevich B, Teixeira MAL, Várbíró G, Ekrem T (2019) DNA barcode reference libraries for the monitoring of aquatic biota in Europe: Gap-analysis

and recommendations for future work. The Science of the Total Environment 678: 499–524. https://doi.org/10.1016/j.scitotenv.2019.04.247

West K, Travers MJ, Stat M, Harvey ES, Richards ZT, DiBattista JD, Newman SJ, Harry A, Skepper CL, Heydenrych M, Bunce M (2021) Large-scale eDNA metabarcoding survey reveals marine biogeographic break and transitions over tropical north-western Australia. Diversity & Distributions 27(10): 1942–1957. https://doi.org/10.1111/ddi.13228

WoRMS Editorial Board (2023) World Register of Marine Species. https://doi.org/10.14284/170 [Accessed 2023-04-05]

Xia X (2020) Improving Phylogenetic Signals of Mitochondrial Genes Using a New Method of Codon Degeneration. Life (Basel, Switzerland) 10(9): 171. https://doi.org/10.3390/life10090171

Xiong F, Shu L, Zeng H, Gan X, He S, Peng Z (2022) Methodology for fish biodiversity monitoring with environmental DNA metabarcoding: The primers, databases and bioinformatic pipelines. Water Biology and Security 1(1): 100007. https://doi.org/10.1016/j.watbs.2022.100007

Zafeiropoulos H, Gargan L, Hintikka S, Pavloudi C, Carlsson J (2021) The Dark mAtteR iNvestigator (DARN) tool: Getting to know the known unknowns in COI amplicon data. Metabarcoding and Metagenomics 3968: e69657. https://doi.org/10.3897/mbmg.5.69657

Zhang S, Zhao J, Yao M (2020) A comprehensive and comparative evaluation of primers for metabarcoding eDNA from fish. Methods in Ecology and Evolution 11(12): 1609–1625. https://doi.org/10.1111/2041-210X.13485

Supplementary material 1

Strings of search motives to query the GenBank accessions annotated descriptions and titles

Authors: João T. Fontes, Kazutaka Katoh, Rui Pires, Pedro Soares, Filipe O. Costa

Data type: pdf

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.8.128646.suppl1

Supplementary material 2

Additional methodology details for conducting species discrimination power analysis using sequences from NCBI GenBank

Authors: João T. Fontes, Kazutaka Katoh, Rui Pires, Pedro Soares, Filipe O. Costa

Data type: pdf

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.8.128646.suppl2

Supplementary material 3

GenBank accession numbers for records utilized in the study

Authors: João T. Fontes, Kazutaka Katoh, Rui Pires, Pedro Soares, Filipe O. Costa

Data type: csv

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.8.128646.suppl3

Supplementary material 4

Number of sequences and species locally aligned for each genomic region

Authors: João T. Fontes, Kazutaka Katoh, Rui Pires, Pedro Soares, Filipe O. Costa

Data type: xlsx

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.8.128646.suppl4

Supplementary material 5

Average congeneric divergence values using mitochondrial genomes

Authors: João T. Fontes, Kazutaka Katoh, Rui Pires, Pedro Soares, Filipe O. Costa

Data type: xlsx

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.8.128646.suppl5

Supplementary material 6

Average congeneric divergence values using independent sequences

Authors: João T. Fontes, Kazutaka Katoh, Rui Pires, Pedro Soares, Filipe O. Costa

Data type: xlsx

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.8.128646.suppl6